

S. M. C. M. L.

```

1  TCACAGAACA  TGTCACAACA  CAGCCCCGAG  TATGCTTTGG  TTTTCACCAT
51  CTGGG3TGCT  ATGGCCACCA  TG3TCTCCAG  TGGCCTGGGT  GCTGCTGTG
101 GCATGGCCAA  GAATGGCACC  GGCATCATGG  CCATGTCTGT  CATGTGGCCA
151 GAGCTGATCC  ACATGAAGTC  CATCATCCCA  GTGGTCATGG  CTGSTATCAT
201 CACCATCTAT  GGCCTAGTGG  CGGCTGTCC  CCGTGCCTAAC  TCCCTGAATG
251 ATGACAACAG  TCTCTATAGC  AGTTTCTCTC  AGCTGGGGCC  TGGCCTGAGT
301 GGCCTGGCAG  CCGGCTTTGC  CATCTCATC  GGGGGGACA  CTGGCAAGTG
351 TGGCACTGCC  CAGCAGCCCC  GACTATTTGT  AGGCATGATA  CTGATCTCA
401 TCTTTGCCAA  GGTGCTCATT  CTCTCCACAA  AGCAGCCCT  CTCAAAACC
451 ACCAGTCACA  GAATAAGATG  TAAAGACCA  CCTCTCAT  TCCGGAACAA
501 ACAGCCTGAC  ACGCATGTGC  TGGGCAGCTG  GCGCCAGTA  GTTGAFTTC
551 TAAGTGTACA  GTGCTCTCT  GTTCATCTC  TTTTGGGAG  GCCTTGGTCC
601 CTCCC3CCCC  ATGCTGTGGA  CATCTGAACC  TAC

```

FEATURES:

5'UTR: 1-9
 Start Codon: 10
 Stop Codon: 625
 3'UTR: 628

HOMOLOGOUS PROTEINS:

Top BLAST Hits:

	Score	E
gi 4562313 ref NP_001685.1 ATPase, H+ transporting, lysosomal ...	181	5e-45
gi 137477 sp P23956 VATL_BOVIN VACUOLAR ATP SYNTHASE 16 KD PROT...	180	8e-45
gi 227919 prf 1713409A H ATPase 16K [Bos taurus]	180	1e-44
gi 3024312 sp O18882 VATL_SHEEP VACUOLAR ATP SYNTHASE 16 KD PRO...	178	3e-44
gi 418179 sp Q03105 VATL_TORMA VACUOLAR ATP SYNTHASE 16 KD PROT...	175	5e-43
gi 5753144 ref NP_033859.1 ATPase-like vacuolar proton channel...	174	8e-43
gi 57554 pir PXBOV6 H+-transporting ATPase (EC 3.6.1.35), vacu...	165	3e-40
gi 137478 sp P23380 VATL_DROME VACUOLAR ATP SYNTHASE 16 KDa PRO...	157	1e-37
gi 3334403 sp O16110 VATL_AEDAE VACUOLAR ATP SYNTHASE 16 KD PRO...	156	2e-37
gi 1718005 sp P55277 VATL_HELVI VACUOLAR ATP SYNTHASE 16 KD PRO...	153	1e-36
gi 401334 sp P31403 VATL_MANSE VACUOLAR ATP SYNTHASE 16 KD PROT...	153	1e-36
gi 10442628 gb AAG17394.1 AF277150_1 (AF277150) V-ATPase 16 kD ...	151	5e-36
gi 7294725 gb AAF50062.1 (AE003544) CG7547 gene product [Droso...	151	7e-36
gi 2493142 sp Q26250 VATL_NEPNO VACUOLAR ATP SYNTHASE 16 KD PRO...	150	2e-36
gi 251354 gb AAB22509.1 vacuolar H(+)-ATPase proteolipid subun...	150	2e-36
gi 249143 sp Q00607 VATL_CANTE VACUOLAR ATP SYNTHASE 16 KD PRO...	144	9e-34

BLAST to dbEST:

	Score	E
gi 9336427 /dataset=dbest /taxon=960...	165	1e-33
gi 9359905 /dataset=dbest /taxon=9606 ...	165	1e-33
gi 9134224 /dataset=dbest /taxon=9606...	165	1e-33
gi 10219114 /dataset=dbest /taxon=96...	165	1e-33
gi 9347217 /dataset=dbest /taxon=960...	165	1e-38
gi 9152104 /dataset=dbest /taxon=9606...	161	2e-37
gi 9894156 /dataset=dbest /taxon=960...	155	1e-35

EXPRESSION INFORMATION FOR MODULATORY USE:

library source:

Expression information from BLAST dbEST hits:

gi|9336427 Human uterus
gi|9359905 Human fetal liver
gi|9134224 Human brain
gi|10219114 Human lung
gi|9347217 Human placenta
gi|9152104 Human skin
gi|9894156 Human ovary

Expression information from PCR-based tissue screening panels

Human Bone marrow
Human Brain
Human Colon
Human Fetal Brain
Human Fetal heart
Human Fetal Kidney
Human Fetal liver
Human Heart
Human Kidney
Human Liver
Human Lung
Human Pancreas
Human Placenta
Human Prostate
Human Skeletal Muscle
Human Small Intestine
Human Spleen
Human Testis

1 MSNNSPEYAL VFTISGAMAT MVSSGLGAAC GMAKNGTGIM AMSVMWPELI
 51 HMKSIIPVVM AGIITIYGLV AAVPPANSLN DDNSLYSSFL QLGAGLSGLA
 101 AGFAIVIVGD TGKCGTAQQP RLFVGMILIL IFAKVLILST KQPLSKPTSH
 151 FIRCKDHPSS FRNKQPDTHV LGSWPSVVDL LSVQCPRVHR LLARPCPLPP
 201 HAVDI

FEATURES:

Functional domains and key regions:

[1] PDC000001 PS000001 ASN_GLYCOSYLATIONN-glycosylation site
 35-38 NGTG

[2] PDC000005 PS000005 PKC_PHOSPHO_SITEProtein kinase C phosphorylation site
 Number of matches: 4

1 111-113 TGM
 2 139-141 STK
 3 149-151 SHF
 4 160-162 SFR

[3] PDC000006 PS000006 CK2_PHOSPHO_SITECasein kinase II phosphorylation site
 Number of matches: 2

1 78-81 SLND
 2 176-179 SVVD

[4] PDC000008 PS000008 MYRISTYLN-myristoylation site
 Number of matches: 8

1 16-21 GAMATM
 2 25-30 GLGAAC
 3 27-32 GAACGM
 4 31-36 GMAKNG
 5 68-73 GLVAAV
 6 93-98 GAGLSG
 7 98-103 GLAAGF
 8 172-177 GSWPSV

Membrane spanning structure and domains:

Helix	Begin	End	Score	Certainty
1	14	34	1.389	Certain
2	37	57	0.733	Putative
3	60	80	2.030	Certain
4	95	115	1.775	Certain
5	127	147	1.699	Certain

BLAST Alignment to Top Hit:

```
>gi|4502313|ref|NP_001685.1| ATPase, H+ transporting, lysosomal
(vacuolar proton pump) 16kD
>gi|137479|sp|P27449|VATL_HUMAN VACUOLAR ATP SYNTHASE 16
KD PROTEOLIPID SUBUNIT >gi|107394|pir||A39367
H+-transporting ATPase (EC 3.6.1.35) chain PKD1 - human
>gi|189676|gb|AAA60039.1| (M62762) vacuolar H+ ATPase
proton channel subunit [Homo sapiens]
Length = 155
```

Score = 181 bits (455), Expect = 5e-45

Identities = 110/153 (71%), Positives = 114/153 (73%), Gaps = 14/153 (9%)

```
Query: 2   SNNSPEYALVFTISGAMATMVSSGLGAACGMAFNGTGIMAMSVMWPELIHMKSIIPVVMA 61
          S + PEYA  F + GA A MV S LGAA G AK+GTGI AMSVM PE I MKSIIPVVMA
Sbjct: 4   SKSGPEYASFFAVMGASAAMVFSALGAAYGTAFSGTGIAAMSVMRPEQI-MKSIIPVVMA 62

Query: 62  GIITIYGLVAAPPPANSLNDDNSLYSSFLQLGA----GLSGLAAGFAIVIVGDTGKCGTA 117
          GII IYGLV AV  ANSLNDD SLY SFLQLGA      GLSGLAAGFAI IVGD G  GTA
Sbjct: 63  GIIAIYGLVVAVLIANSLNDDISLYKSFLQLGAGLSVGLSGLAAGFAIGIVGDAGVRGTA 122

Query: 118 QPRLFVGMILILIFAKV-----LILSTK 141
          QPRLFVGMILILIFA+V          LILSTK
Sbjct: 123 QPRLFVGMILILIFAEVLGLYGLIVALILSTK 155
```

Hmmer search results (Pfam):

Model	Description	Score	E-value	N
PF00137	ATP synthase subunit C	14.8	0.028	2

Parsed for domains:

Model	Domain	seq-f	seq-t	hmm-f	hmm-t	score	E-value
PF00137	1/2	12	73 ..	1	65 [.	7.6	2.4
PF00137	2/2	89	133 ..	1	53 [.	14.6	0.031

1	GCTGTGGGGC	CAGGAAAAGG	AGAGAAGGTG	AAAACCCCGT	CAGTCCCTCA
51	CAATCAGCAC	GTGGAAATCT	AGAAATGCAG	GAGAGGCGTG	CACTCATGGT
101	GGAATCCAGA	ATGAAAGAGG	TGGACGACTG	AATGAGCAGA	AGGAGGCAAG
151	CACCAGAGGC	TTGGGGGTCA	CATTTCTTGG	AAGTGGCGTG	GAGCTGGCAG
201	ATGAGAACTC	TGGCTACCGG	TCTTCATTCC	ACTAACAGTA	GCTCCTCTAA
251	CGACATGCCC	CTTCCCTCTG	TACCCCGCTC	CGCATGCGGC	AAGTAGTTCC
301	CGGACGGGAC	CCTTCCCCCT	GTACCCCGCT	CGCATGGGG	CCAGTAGTTC
351	CCGGACGGGC	CCCTTCCCTC	TGTACCCCGC	TCCGCATGCG	GCAAGTAGTT
401	CCCGGACGGG	CCCCTTCCCT	CTGTACCCGG	CTCCGCATGC	GGAAAGTAGT
451	TCTACGGTG	TGGTTTTTGC	ATGTAGATGA	AACCTTTTGA	GGGATAAAGG
501	TTTTTTTTTT	AAGTACTTTA	GCAAAATGCA	ACTGTTATTA	TCAATATTAG
551	CCAGCATCTT	TTTTTTTTTT	TTTTTTTTTT	TTTTTGASAT	GGAGTTTGGC
601	TCTTGTACCC	CAGGCTGGAG	TGCAATGGGA	AAATCTAGGC	TCACTGCAAC
651	CTCCCGCTCC	CAAGTTCAAG	CGATTCTCCT	CCCTCAGGCT	CCCAGGTAGC
701	TGGATTACA	GGCGTGTGCA	ACCAACGCCA	CTAATTTTTT	GTATTTTTAG
751	TAGAGACAGG	GTTTCACCAT	GTTGGCCAGG	CTGCTCTCGA	ACTCCTGACC
801	TGATGTGATC	CATCCGCCTC	AGCCTCCCAA	AGTCTGCGGA	TTACGTAGCC
851	AGTGTCTTTC	TTAAGTGCCT	GTCAAAATAAT	GCTCCTGGTT	TATAAGTCCC
901	CCTGGCTCTA	CCTTCTGGGT	GCTCAGACAC	CAACACAGAG	AGAACAGAAAT
951	TAACTCCTCG	AGAAGTTACA	TATCTAATAA	TATAAAGAGT	AAGATTGTGA
1001	GGAAACTGCA	GGGGAAGCAG	GTAAGTTAGG	AAAAAGTATC	CTCACITTTT
1051	TGCTGACCGA	TGAGTCATAA	TTCTTGAAAT	TGGTGCTGG	AAAGGTCCAT
1101	TAAATATTCC	AGGAGATTCT	AGGGAAGTTC	CAGAAATGGTA	GAAGAACTGG
1151	AACCATAAAG	CCTGGGGAAG	GGATGGAAGT	CTTGGGAAA	GAAGCACTAA
1201	ACAGCCCACT	GGAGACAAGG	AAAGACTGGT	CTCTCTGTG	CTTCGAGCCC
1251	AGCAATGATT	ATTCACTCAG	ATATGCCCGG	GCAGCTCTTG	CTGCTAGAGC
1301	CAATGTCTTT	CCCAGACCCA	GGCAAGGTTG	CATCCTACCC	CTGACAGGAA
1351	ACAGGCGCAG	AGGTGGGGCT	GCCCCGGGTG	CTTGGTGTG	GGAGGGGGCG
1401	GGGGGAATCC	CGGGTGTGG	GAGGACAAGG	CAGAGTCAGC	TAGCTGTGAG
1451	GCTAGGGGAG	AAGACCTCTC	TAGTCTGGGA	GAGACCCCTC	CTTTCCCTAG
1501	TCTTTGTAAT	TCCAAAAAGG	CAGGCTTCCCT	GTGTTTACTA	ACCATAACAG
1551	GACTGACTAT	ACAGCAGCCA	GAAAGATTCT	GAGAAACCTG	TGATAGAGAA
1601	AAACAGATGC	GGAAGCGGGA	GAAGAGAATT	TCATAGGACA	CTAGGGAAAG
1651	AGAAATGGAA	TCTGTGGTCT	AAAGAGGGAA	CCAAAGTCTGG	CCAAACATGGT
1701	GAAACCCCAT	CTCTATTACA	AATACAAAAA	TTAGCTGGGC	ATGGTAGTGC
1751	ATGCTGTAA	TCCAGCTTAC	TCAGGAAGGT	AAGGCATGAG	AATCACTTGA
1801	GACTGGGAAG	CGGAGGTTGC	AGTGAAGCCA	GACTGCACCA	CTGCACTCCA
1851	GCTGGGCAA	CAGAGCAAGA	CCTCGTCTCA	AAAAAAAAAA	AAATTTAAAT
1901	TTAAATTTAA	AAAAAATAAA	CAGGGAACCA	ACAAAGAGTG	GCAGAACAGA
1951	ATAAAGTCTC	AAGCCAAATA	ACTCCCTTGC	CTTGGAAAGAA	CAAGGCTGCC
2001	AGCTCTTTGG	AGCCTCTGTT	TATCGGCTAC	CAATTCAAAG	GACAGTGAGC
2051	CTGAGCTGGC	CTGGGAGGCC	CTCCCCCTC	CCAGATGAAA	ACAATAGGCC
2101	TGTTTTCTTG	AGCTCTTCCCT	GTAATCCAGA	AGGCCACAC	AGAGAGGAAAG
2151	AGGGGGGCAA	AGGCAGTGGC	TATACCCAGT	GGGGGAGGGG	ATATTTAGCC
2201	TCCCATAAAT	TCATCAGCTC	CCTTAAAGAC	ACCCCAAAAC	ACCAACAATC
2251	TAAATGTATA	AATAGTGACT	GCTATGCAAA	TGGAAGTTTA	AAACCTATCC
2301	CTTAGCCCCA	GTACCACAG	ATTACTAAC	CTAAACCCCA	TCTGTAGGAG
2351	ATATTCTTAA	GCCACCACAG	GGAAGGGGAT	AAGGGCTTGA	GAGACAAAGG
2401	ACAGATGGGG	TCTCCCCAAC	AATTTAAGTT	AAGTTCCACA	AGGATACAGT
2451	ACTGGTACAG	ATTTGGAAGT	AGGGGCAAGT	ATTCTGACAG	AAGGGTGGTG
2501	TCTTAGCCAC	CCTTCAATTA	GAGTAGCTA	AAGGCTGTGT	GTGTGTCTGT
2551	GTGTGTGCAT	AAGAAAAGAA	ATAGGAGGGT	GTGTGTGTGG	TAAGAAAAGG
2601	CATCTTGGCT	GGCGCGGTG	GCTACCCCT	ATAATGCGAG	CACTTTGGGA
2651	TGGGAAGGCT	GGCGGATTGC	CTGAGCTCAG	GAGTTTGAGA	CCATACGGGG
2701	CAACATGGTG	AAACCCCATC	TCTACTAAAA	ATACAAAAAA	TTAGCTGGGC
2751	ATGGGGTGGC	GTGCCATATG	TTCCAGCTAC	TGGGGAAGGT	GAGGCATGAG
2801	AATGGCTTGA	GCCTTGAGAG	CAGAGGTTGA	AGTGAAGCTG	GATCGCACCA
2851	TGCTATTCCA	GCTTGGGCTA	CAGAGTGACA	CTCCATCTCA	AAAAAAAAAA
2901	AAAAAAAAAA	AAAAACCAGC	ATCTTTGCTG	CCACTAGTCC	ACTGTCTTGG
2951	CATCTACTCT	CTGCCATGCC	CATGCTTGTG	CCGCTCCCCA	CTCACAGACA
3001	TGTCGAACAA	CAGCCCCGAG	TATGCTTTGG	TTTTTACCAT	CTCGGGTGGT
3051	ATGGCCACCA	TGGTCTCCAG	TGGCCTGGGT	GCTGCCTGTG	GCATGGCCAA
3101	GAATGGCACC	GGCATCATGG	CCATGTCTGT	CATGTGGCCA	GAGCTGATCC

FIGURE 3, page 1 of 4

3151	ACATGAAGTC	CATCATCCCA	GTGGTCAATGG	CTGGTATCAT	CACCATCTAT
3201	GGCCTAGTGG	CGGCTGTCCC	CCCTGCCCAC	TCCCTGAATG	ATGACAACAG
3251	TCTCTATAGC	AGTTTTCCTCC	AGCTGGGGCC	TGGCCTGAGT	GGCCTGGCAG
3301	CCGGCTTTGC	CATCGTCATC	GTGGGGGACA	CTGGCAAGTG	TGGCACTGCC
3351	CAGCAGCCCC	GACTATTTGT	AGGCATGATA	CTGATCCTCA	TCTTTGCCAA
3401	GGTGTTCATT	CTCTCCACAA	AACAAGCCCT	CTCAAAAACC	ACCAGTCACA
3451	GAATACGATG	TAAAGACCAC	CCCTCCCTCAT	TCCGGAACAA	ACAGCCTGAC
3501	ACGCATGTGC	TGGGCAGCTG	GCCCTCAGTA	GTTGATCTTC	TAAGTGTACA
3551	GTGTCTCTGT	GTTTCATCGTC	TGTTGGCCAG	GCCTTGGCCC	CTCCCGCCCT
3601	ATGCTGTGGA	CATCTGAACC	TACTCATCAC	CCATCCAGGT	CCCCGACCAG
3651	TGAGGACTCA	GGCCCCCTGGA	TGCCCCACCC	ATCTCCCTTG	AGTACTCTAT
3701	GTATAAGGAT	GAATTAGAGT	TGTCATTTTC	TCTTCATTAG	ATATTTATAA
3751	AATTTTGGCC	TGTCATACCC	CCCTGTGGAG	AGCCCTCATC	TCCCACCTAT
3801	CTGTCACTGC	ATGGAGGTTT	CCATTGCGGA	GGCTCCTTTG	ATGGAACCA
3851	CTCTCCACGC	CCGCGCTGCC	AGGCCCTGCG	CGGCAGCTGT	GTCTGATAAA
3901	GTTCTCAGAT	GTCGGGGGAG	GGAAAAGAAA	AAAAAAGAGA	GTGTGAGTAC
3951	GTAAAGAGAGA	GAAAACGGGA	GTGGGTGTGT	GAGCTGGAGA	CAGGGAACCT
4001	GCAGGAAAAG	TCTGATAAGA	TCACTCTCTT	CTTACCCCAAG	CAGAGATACT
4051	GGACACAGCC	CCTCAAGGAC	CCAGAGGCTA	AGTACAGGCG	GAGATGCTTG
4101	CTTTTCTCAA	TGGGAGGTGG	CCTCCACGGC	CTGAGGAAGT	CTCCATTATC
4151	CCAGAGGACA	ACTAGGAAGC	AGGTAGACAG	CATCATCCCC	ACTTATACCC
4201	CAAGGTGCTT	GGGGTGAATG	GCAAGGCCAA	AGCCAAAGCA	TGAGACAGAT
4251	TAAATGTTC	TATGGCGAGA	GAAAGAGAA	GCTTCACCCAG	CATCTCTCCA
4301	CTGAGCAAA	GAAAGGAAGA	GAGAAAGCAG	GCTGATACCC	TCATCAATTT
4351	CTTACTGTTC	ATGATATACC	ACCATCAACT	GGACTTTTTT	TTTTTTTTTT
4401	AGATAGAGTC	TGCTTTTTGT	CACCCAGGCT	GGAGTGCACT	GGCATGATCT
4451	CAGCTCACTG	CAACTTCCAT	CTCCAGGTTT	CAAATGATTG	TCCCGCCTCA
4501	GCCTCCTGAG	TAGCTGGGAT	TACAGGTGCC	TGCTACCCACA	TCCAGCTGAT
4551	TTTTTTTGT	TTTTTAGTAG	AGATGCGGTT	TCTTTCTTTT	TTTTTTTTTT
4601	TTTTGAGACG	GATCTTTGCT	CTGTCCGCCA	GGCTGGAGTG	CAGTGGCGCG
4651	ATCTCGGCTC	ACTGCAACCT	CCGCTCCCCA	GCTTCAAGCC	ATTCTCCTGC
4701	CTCAGCCTCC	CGAGTAGCTG	GGACTACAGG	CACCTGCCAC	CACACTCGGC
4751	TAAATTTTTG	TATGTTTAGT	AGATATGGGG	TTTCACTGCT	GTCTCAACCT
4801	TCTACCTCA	TGATCCGCCC	GCCTCGGCTT	CCCAAAGTGC	TGGGATTACA
4851	GGCATGAGCC	ACTGTGCCCG	GCCTTTTTTT	TTTTTTTTTG	GATGGAGTCT
4901	CTCTCTGTGG	CCCAGGCTGG	AGTGCAATGC	CACAATCTCA	GCTCACTGCA
4951	AGCTCCACCT	CCGAGGTTCA	CGCCATTCTC	CTGCTCAGC	CTCCTGAGTA
5001	GCTGGGACTA	CAGGCGCCCG	CCACCACGCC	CAGCTAATTT	TTTGTATTTT
5051	TAGTAGAGAC	GGGGTTTCAC	CTTGTTAGCC	AGGATGGTCT	TGATCTCCTG
5101	ACCTGCTGAT	CCACCTGCC	CAGCCTCCCA	AAGTGCTGGG	ATTACAGGTG
5151	TGAGCCACCA	TGCTGGCCTT	TTTTTTTTTT	TTTAAGACAG	GAGTGTGGTG
5201	GCAACATCTC	AGCTCACTGC	AACCTCCCTT	TCTAGGTTCA	AGCAATTCTC
5251	CTGCTCAAGC	TTCCTAAGTA	TAGTAATAGC	TGGGACTATA	GGCGCCCACT
5301	ACCAAGCCCG	GCTAATCTTT	TGTATTTTTA	GTAGAGATGG	GGTTTCAACA
5351	TGTTGGCCAG	GCTGGTCTCG	AATTGCTGAC	CTCAAGTGAT	CTGCCCACCT
5401	GGGCTCTCCA	AAGTGCTGGG	ACTATAGGCG	GGAGCCACCG	CGCCAGGCTT
5451	GGAATCTTTT	TAATGAGGCC	TTCAAAAAAA	CTCCTTTTCT	CAGCGCTTCT
5501	TACTCTCTGA	AACAGACTCT	CCACTCTGCT	AACCTGCCTT	CTCACACTGT
5551	GGAACCTCAA	CGGATCTTTT	TATTCTGAAT	CCACAACGTG	AAGTACTTGT
5601	CTCTCTCTTA	TGGATGGCTA	CCTGTGTTTT	GAAGTGTTTT	TATGGGAATG
5651	AACTACTGGA	GGGGAGGAAA	TCAAGCCAGT	TCTAGAAGTA	GAAAGGAAGG
5701	GAAAGAAACA	GAAAAAATAT	TTATGTGATG	GGAGGAAAGG	CAGTTTATAA
5751	ATCACTCATG	GATCTCTATG	CCAGAGGGAT	GTGTGAGACA	CACGCATGCA
5801	CACACACACT	GACTTGCAGG	TACATGCAGA	GGCAGAAACA	AGTCAGGACA
5851	TGACACATAA	ATGAATACAA	ATACCATTCT	CATCAGAAAC	CAGTCAGAGC
5901	AGAGGGGCCC	TGCTGGAGCC	AAGGAGACTG	GAATTTATTC	CCCTCCTCCT
5951	CTCAAAAGGT	AATTTTGTCT	CCTCATGTCT	TAGGTTCCCC	ACAGATCTGG
6001	CTGCTCAGAA	CAGGGGCCCT	GCTCTGCTGG	CTGACTCAG	CCTGGAGGTC
6051	TTCAAGATG	GAGGCTTATA	AGAAGTGGCA	GCTGACACCT	GGAGGGAGCT
6101	GGATGAAAAG	AGGCAGTGCA	GAGTAGAGAA	AGCCAGGTGG	TGGGGGAGGG
6151	AGTGAAGGAG	AAGAGGGGAC	CAGATTCAAG	CAGCCTTGCG	CTGGTTCTAA
6201	AATGGCCACA	GCAAGGCAAC	GGACAGATGG	TCCCTTTCTG	ATGCTGAGCC
6251	GGGGAAGTGG	GGAAAGGGAA	AAGGAAAAAA	TAAACACCAT	CACAGTCAGA

FIGURE 3, page 2 of 4

6301 AATTTAAAAA TAAACTGAAA AACCTAAAAA ATAAACCGT

FEATURES:

Start: 3000
Exon: 3000-3614
Stop: 3615

CHROMOSOME MAP POSITION:

Bac accession number: AC005973
Chromosome 17

ALLELIC VARIANTS (SNPs):

DNA			
Position	Major	Minor	Domain
559	-	A T	Beyond ORF (5')
3638	G	C	Beyond ORF (3')
5446	C	T A	Beyond ORF (3')
5808	A	G	Beyond ORF (3')
5892	A	C	Beyond ORF (3')
6071	A	G	Beyond ORF (3')

Context:

DNA

Position

559 CCCTTCCCTCTGTACCCCTCTCGCATCTGGCAAGTAGTTCCCGGACGGCGAACCCTTCCCT
CTGTACCCCGCTCCGCATGGGGCCAGTAGTTCCCGGACGGCGCCCTTCCCTCTGTACCCCT
GCTCCGCATGGGGCAAGTAGTTCCCGGACGGCGCCCTTCCCTCTGTACCCCGGCTCCGCAT
GGGAAAGTAGTTCTACGGTCTTGGTTTTGCATGTAGATGAAACCCCTTTGAGGGGTAA
GGTTTTTTTTTAAAGTACTTTAGCAAAATGCAAACTGTTATTATCAATATTAGCCAGCATC
[-,A,T]
TTTTTTTTTTTTTTTTTTTTTTTTTTTTTGGAGATGGAGTTTCGCTCTTGTCAACCAGGCTGGA
GTGCAATGGCAAAATCTAGGCTCACTGCAACCTCCGCCTCCCAAGTTCAAGGGATCTCC
TGGCTCAGCCTCCAGGTAGCTGGGATTACAGGGGTGTGCAACCACACCCCACTAATTTT
TGTATTTTATGTAGAGACAGGGTTTCAACCATGTTGGCCAGGCTGGTCTCGAACTCCTGAC
CTCATGTGATCCATCCGCCTCAGGCTCCCAAGTGTGGGATTACGTAGCCAGTGTCTTT

3638 GTGTGGCACTGCCCAGCAGCCCCGAGTATTTGTAGGCATGATACTGATCCTCATCTTTGC
CAAGGTGCTCATTCTCTCCACAAAGCAGCCCTCTCAAAACCCACCAGTCAAGAAATAG
ATGTAAAGACCAACCTCTCTATTCCGGAACAAACAGCCTGACACGCATGTGCTGGGCAG
CTGGCCCTCAGTAGTTGATCTTTCTAAGTGTACAGTGTCTCGTGTTCATGGTCTGTGGC
CAGGCCCTTGGCCCTCCCGCTCCATGCTGTGGACATCTGAACCTACTCATCAACCATCCA
[G,C]
GTCCCCGACCAAGTGAGGACTCAVGGCCCTGGATGCCCCACCCATCTCCCTTGAATCTCT
ATGTATAAGGATGAATTAGAGTTGTCTATTTCTCTTATTAGATATTTATAAAGATTGG
CCTGTCCATACCCCTGTGGAGCAGCCCTCATCTCCACCTATCTGTCACTCATGGAGGT
TCCCATTGGGGAGGCTCCTTGGATGGAAACCCCTCTCCAGCCCGCGCTGGCAGGCCCTG
TGCGGCAGCTGTGTCTGATAAAGTTCTCAGATGTGCGGGGAGGGAAAGAAAAAAGGA

5446 AGGTGTGAGCCACCATGCTGGCCTTTTTTTTTTTTTTAAGACAGGAGTGTGGTGGCACA
ATCTCAGCTCACTGCAACCTCCCTTCTAGGTTCAAGCAATTCTCCTGCTCAGCTCTCT
AAGTATAGTAATAGCTGGGACTATAAGGGCCACCACACGCCCCGCTAATCTTTTGTAT
TTTTATAGAGATGGGTTTCAACCATGTTGGCCAGGCTGGTCTCGAATTGCTGACCTCAA
GTGATCTGCCACCTGGGCTCCCAAAATGCTGGGACTATAGGCGGGAGCCACCGGGCC
[C,T,A]
GCCTGGACTCTTTTTAATGAAGCCTTCAAAAAAACTCCTTTTCTCAGCGCTTCTTACTCT
CTGAAACAGACTCTCCACTCTCTTAAGCCCTGCTCTCACACTGTGGAACCAACGGGATC
TTTTTATTCTGAATCCCAACCTGGAAGTACTTGTCTCTGTCTATCGATGGCTACCTGTG
TTTTGAAGTGTTTTTATGGGAATGAAGCACTGGAGGGGAGGAAATCAGGCCAGTTCTAGA
AGTAGAAGGAAGGCGAAGAAACCAAGAAAAATATTTATGTGATGGGAGGAAAAGCACTTT

5808 TGAAACAGACTCTCCACTCTGCTAACCCCTGCCTCTCACACTGTGGAACTCAACCGGATCT
TTTTATTCTGAATCCACAACTGTGAAGTACTTGTCTCTGTCTATCGATGGCTACCTGTGT
TTTGAACTGTTTTTATGGGAATGAAGCACTGGAGGGGAGGAAATCAGGCTAGTTCTAGAA
GTAGAAAGAAAGGGAAGAAACAGGAAAAATATTTATGTGATGGGAGGAAAGGCAGTTTA
TAAATCACTCATGGATCTCTATGCCAGAGGGATGTGTGAGACACACGCATGCCACACAC
[A, G]
CTGACTTGCAGGTACATGCAGAGGCAGAAACAAGTCAGGACATGACACATACATGAATAC
ACATACCATTTCTCATCAAAAACAGTCAGAGCAGAGGGGGCCCTGCCTGGAGCAAGGAGAC
TGGAAATTTATTCCCTCTCTCTCTCAAAGGGTAATTTTGTCTGCCTCCATGTCTAGGTTCC
CCACAGATCTGGCTGCCTCAGACAGGGGGCCCTGGTCTGGTGGCTGGACTCAGCCTGGAGG
TCTTTCACAGATGGAGGCTATAAGAGGTGGCAGCTGACACCTGGAGGGAGCTGGATGAAA

5892 AGTACTTGTCTCTGTCTATCGATGGCTAOCCTGTGTTTTTGAAGTGTTTTTATGGGAATGA
AGCACTGGAGGGGAGGAAATCAGGCTAGTTCTAGAAGTAGAAGGAAGGGGGAAGAAACAG
GAAAAATATTTATGTGATGGGAAGGAGGAGGCTTTATAAATCACTCATGGATCTCTATGC
CAGAGGGATGTGTGAGACACACGCATGCCACACACACTGACTTGCAGGTACATGCAGAG
GCAGAAACAAGTCAGGACATGACACATACATGAATACACATACCATTTCTCATCAAAAAC
[A, C]
GTCAGAGCAGAGGGGGCCCTGCCTGGAGCAAGGAGACTGGAATTTATTCCCTCTCTCTCT
CAAAGGGTAATTTTGTCTGCCTCCATGTCTAGGTTCCCCACAGATCTGGCTGCCTCAGACA
GGGGCCCTGGTCTGGTGGCTGGACTCAGCCTGGAGGTCTTACAGATGGAGGGCTATAAG
AGGTGGCAGCTGACACCTGGAGGGAGCTGGATGAAAGCAGGCAGTGCAGAGTAAGAGAAAG
CCAAGTGGTGGGGAGGGAGTGGAGGAGAGAGAGGGGACCAGATTCAAGCAGCCTTGGCT

6071 CCAGAGGGATGTGTGAGACACACGCATGCCACACACACTGACTTGCAGGTACATGCAGAG
GGCAGAAACAAGTCAGGACATGACACATACATGAATACACATACCATTTCTCATCAAAAAC
CACTCAGAGCAGAGGGGGCCCTGCCTGGAGCAAGGAGACTGGAATTTATTCCCTCTCTCT
CTCAAAGGGTAATTTTGTCTGCCTCCATGTCTAGGTTCCCCACAGATCTGGCTGCCTCAGAG
CAGGGGCCCTGGTCTGGTGGCTGGACTCAGCCTGGAGGTCTTACAGATGGAGGGCTATA
[A, G]
GAGGTGGCAGCTGACACCTGGAGGGAGCTGGATGAAAGCAGGCAGTGCAGAGTAAGAGAA
GCCAGGTGGTGGGGGAGGGAGTGGAGGAGAGAGAGGGGACCAGATTCAAGCAGCCTTGGCT
TGGTTCTAAAATGGCCACAGCAAGGCAACGGACAGATGGTCCCTTTCTGATGCTGAGGCG
GGGAAGTGGGGAAAGGGAAAAAGAAAAAATAAACCATCACAGTCAGAAATTTAAAAAT
AAACTGAAAAACCTAAAAAATAAACCT